Statistical Methods and Scientific Inference

by

Sir Ronald A. Fisher, Sc.D., F.R.S.

D.Sc. (Adelaide, Ames, Chicago, Harvard, Indian Statistical Institute, Leeds, London)
LL.D. (Calcutta, Glasgow)

Honorary Research Fellow, Division of Mathematical Statistics, C.S.I.R.O., University of Adelaide; Foreign Associate, United States National Academy of Sciences; Foreign Honorary Member, American Academy of Arts and Sciences; Foreign Member, American Philosophical Society; Honorary Member, American Statistical Association; Honorary President International Statistical Institute; Foreign Member, Royal Swedish Academy of Sciences; Member, Royal Danish Academy of Sciences; Member, Pontifical Academy; Member, Imperial German Academy of Natural Science; formerly Fellow of Gonville and Caius College, Cambridge; formerly Galton Professor, University of London; and formerly Balfour Professor of Genetics, University of Cambridge.

Hafner Press
A Division of Macmillan Publishing Co., Inc.
New York
Collier Macmillan Publishers
London

odds are often claimed, evidently with a view to raising the resistance felt to accepting what is intrinsically improbable to such a pitch that conclusion (*d*), although itself repugnant, shall be accepted in preference. The incredulous, however, tend to prefer explanations of types (*a*), (*b*) or (*c*) either to accepting such a claim as, let us say, "precognition", or, what seems almost always to be the last choice, to the acceptance as genuine of a very rare contingency.

The fact, important for the understanding of logical situations of this kind, that reluctance to accept a hypothesis strongly contradicted by a test of significance is not removed, though it may be outweighed, by information *a priori*, is exhibited also by the consideration that if the proposed datum, "The odds are a million to one *a priori* that the stars should really be distributed singly and at random"—if this datum were considered as a *hypothesis*, it would be rejected at once by the observations at a level of significance almost as great as the hypothesis, "The stars are really distributed at random", was rejected in the first instance. Were such a conflict of evidence, as has here been imagined under discussion, not in a mathematical department, but in a scientific laboratory, it would, I suggest, be some prior assumption, corresponding to an axiom or a datum in a mathematical argument, that would certainly be impugned.

The attempts that have been made to explain the cogency of tests of significance in scientific research, by reference to supposed frequencies of possible statements, based on them, being right or wrong, thus seem to miss the essential nature of such tests. A man

who "rejects" a hypothesis provisionally, as a matter of habitual practice, when the significance is at the 1% level or higher, will certainly be mistaken in not more than 1% of such decisions. For when the hypothesis is correct he will be mistaken in just 1% of these cases, and when it is incorrect he will never be mistaken in rejection. This inequality statement can therefore be made. However, the calculation is absurdly academic, for in fact no scientific worker has a fixed level of significance at which from year to year, and in all circumstances, he rejects hypotheses; he rather gives his mind to each particular case in the light of his evidence and his ideas. It should not be forgotten that the cases chosen for applying a test are manifestly a highly selected set, and that the conditions of selection cannot be specified even for a single worker; nor that in the argument used it would clearly be illegitimate for one to choose the actual level of significance indicated by a particular trial as though it were his lifelong habit to use just this level. Further, the calculation is based solely on a hypothesis, which, in the light of the evidence, is often not believed to be true at all, so that the actual probability of erroneous decision, supposing such a phrase to have any meaning, may be, for this reason only, much less than the frequency specifying the level of significance. A test of significance contains no criterion for "accepting" a hypothesis. According to circumstances it may or may not influence its acceptability.

On the whole the ideas (*a*) that a test of significance must be regarded as one of a series of similar tests applied to a succession of similar bodies of data, and (*b*) that the purpose of the test is to discriminate or